



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira de Bejaia
Faculté des Sciences Exactes
Département d'Informatique

ECOLE DOCTORALE RESEAUX ET SYSTEMES DISTRIBUES

Mémoire de Magistère

En Informatique

Option : Réseaux et Systèmes Distribués

Thème

**Tagging Collaboratif et Filtrage de Tags à base du Profil
Utilisateur**

Présenté par

Mme DAHAK-KICHOU Saida

Devant le jury composé de :

Président	TARI Abdelkamel	M.C.A	Université de Bejaia
Rapporteur	AMGHAR Youssef	Professeur	INSA Lyon France
Examineur	DRIAS Habiba	Professeur	USTHB Alger
Examineur	MAHDAOUI Latifa	M.C.A	USTHB Alger
Invitée	MELLAH Hakima	Chargée de recherche	CERIST Alger

Promotion : 2008/2009

Remerciements

Merci au bon Dieu, le tout Puissant.

Je tiens à remercier vivement mon directeur de thèse Mr Amghar Youssef, professeur à l'INSA, pour ses corrections, conseils et orientations. Merci d'avoir accepté de m'encadrer et me guider dans le monde si vaste de la recherche. Qu'il trouve en ce mémoire l'expression de mon profond respect.

Je remercie tout particulièrement Mme Mellah Hakjma, chargée de recherche au CERIST, codirecteur de thèse, de m'avoir orienté, corrigé mon travail, soutenu et encouragé. Merci pour sa disponibilité et sa gentillesse sans égale. Je ne la remercierai jamais assez, qu'elle trouve en ce mémoire l'expression de ma profonde gratitude et mon respect infini.

Mes vifs remerciements et respects s'adressent aux membres du jury Mr Tari Abdelkamel, maitre de conférence à l'université de Bejaia, Mme Drias Habiba, professeur à l'USTHB et Mme Mahdaoui Latifa, maitre de conférence à L'USTHB pour m'avoir fait l'honneur d'accepter de juger ce modeste travail.

Je tiens également à remercier Mr Badache Nadjib, professeur à l'USTHB et Directeur du CERIST, de m'avoir donné la chance d'entrer dans le monde de la recherche.

Mes sincères remerciements vont également à Mr Meziane Abdelkrim, chargé de recherche et responsable de la division systèmes d'information et systèmes multimédias pour ses encouragements et sa compréhension.

Je remercie infiniment Mr Dahak Fouad, maitre assistant à l'école supérieure d'informatique pour son aide très précieuse, ainsi que Mr Azouaou Faïçal, maitre de conférence pour ses orientations et conseils.

Mes chaleureux remerciements vont à mes très chers parents qui ne cessent de croire en moi et m'encourager. Merci pour toute ma famille, grands parents, oncles et tantes, cousins et cousines, et toute ma belle famille.

Mes chaleureux remerciements vont également à mes amis du CERIST Nour ElHouda, Linda, Lamia, Fouzia, Safia, Amel et Leila pour leurs soutiens et encouragements. Merci à tous mes collègues de la division Imen, Badiaa, F.Zohra, Lamia, Djalila, Lydia, Salem, Kamal et Nouredine pour leurs encouragements.

Je remercie mes collègues de la promotion, Lamia, Sahar, Dalila, Faiza, Nabil, Madjid et Rafik pour la bonne ambiance vécue tout au long d'une année de cette belle expérience.

Merci à Mme Zaidi et tous les collègues du service formation pour leur disponibilité.

Je remercie mes amis Malika.B, Malika.A, Radia, Sekoura, Soraya et Fazia pour le soutien moral tant apprécié.

Fouad, merci encore une fois, je ne te remercierai jamais assez pour tout ce que tu fais pour moi, merci d'être à mes côtés dans les moments difficiles, sans toi je n'aurais jamais pu tenir jusqu'à la fin.

Tout simplement Merci !

A mes très chers parents Rahima et Abdelmadjid

A mon mari Fouad

A mes deux adorables petits anges Anis et Yacine

A ma chère sœur Samia

A mes frères Mohamed et Azedine

A mes beaux parents Said et Wardia

A jida Adidi, jeddi Omar et ami Mustapha

A toute ma famille et belle famille

RESUME

Le 'Tagging collaboratif' ne cesse de gagner une popularité sur le web 2.0, cette nouvelle génération du web qui fait de l'utilisateur un lecteur-rédacteur. Le 'Tagging' est un des moyens permettant à l'utilisateur de s'exprimer librement via des ajouts d'étiquettes appelées 'Tags' à des ressources partagées. L'un des problèmes rencontrés dans les systèmes du Tagging actuels est la définition des tags les plus appropriés pour une ressource. Les tags sont généralement classés par ordre de popularité tel que del-icio-us. Or la popularité du tag ne reflète pas toujours son importance et sa représentativité vis-à-vis de la ressource à laquelle il est associé. Partant des hypothèses qu'un même tag pour une ressource peut prendre des significations différentes selon les utilisateurs, et un tag issu d'un utilisateur connaisseur serait plus important qu'un tag issu d'un utilisateur novice, nous proposons une approche de pondération des tags d'une ressource à base du profil utilisateur. Pour ceci nous définissons un modèle utilisateur permettant son intégration dans le calcul du poids d'un tag ainsi qu'une formule de calcul de ce dernier à base de trois facteurs concernant l'utilisateur à savoir, le degré de rapprochement entre ses centres d'intérêts et le domaine de la ressource, son expertise et son estimation personnelle vis-à-vis des tags qu'il associe à la ressource. Un descripteur de ressource contenant les meilleurs tags est ainsi créé.

Mots-clés

Annotation, Tagging collaboratif, profil utilisateur, recherche d'informations.

ABSTRACT

The 'Collaborative Tagging' is gaining popularity on Web 2.0, this new generation of Web which makes user reader/writer. The 'Tagging' is a means for users to express themselves freely through additions of label called 'Tags' to shared resources. One of the problems encountered in current tagging systems is to define the most appropriate tag for a resource. Tags are typically listed in order of popularity, as del-icio-us. But the popularity of the tag does not always reflect its importance and representativeness for the resource to which it is associated. Starting from the assumptions that the same tag for a resource can take different meanings for different users, and a tag from a knowledgeable user would be more important than a tag from a novice user, we propose an approach for weighting resource tags based on user profile. For this we define a user model for its integration in calculating the weight of a tag and a formula for calculating it based on three factors namely the user, the degree of approximation between its centers interest and the field of resource, expertise and personal assessment for tags associated to the resource. A resource descriptor containing the best tags is created.

Keywords

Annotation, Collaborative Tagging, user profile, information retrieval.

ملخص

اكتسب التوسيم التعاوني شعبية كبيرة في الواب 2.0 ، النسخة الجديدة من الشبكة العنكبوتية التي جعلت المستخدم قارئ و محرر في آن واحد. التوسيم أحد الوسائل التي تسمح للمستخدم أن يعبر عن رأيه بكل حرية و ذلك بإضافة علامات بشكل كلمات تسمى 'Tag' لموارد مشتركة على الشبكة. من بين نقائص أنظمة التوسيم التعاوني الحالية التعريف بالعلامات المناسبة أكثر لمورد ما. هذه العلامات مرتبة حسب شعبيتها (عدد مرات ذكرها) مثل النظام Delicious. لكن شعبية العلامة لا تعكس دائما أهميتها و تمثيلها بالنسبة للمورد المنسوبة إليه. بافتراض أنه يمكن لعلامة ما أن تحمل معاني مختلفة حسب المستخدمين، و علامة مقترحة من مستخدم خبير أكثر أهمية من علامة مقترحة من مستخدم مبتدأ، نقترح نظاما لترجيح العلامات مستندا على مواصفات المستخدم. لذلك نقترح نموذجا لمواصفات المستخدم، يسمح هذا النموذج بإدخال المواصفات في عملية ترجيح العلامات. و صيغة لحساب وزن كل علامة مستندا على ثلاثة عوامل: درجة التقارب بين المورد و مراكز اهتمام المستخدم، خبرة المستخدم و تقييمه الشخصي لعلاماته. بذلك نكون مجموعة أحسن العلامات.

الكلمات المفتاحية: الملاحظات، التوسيم التعاوني، مواصفات المستخدم، استرجاع المعلومات.

Table des Matières

Introduction Générale

Introduction	1
Contexte du travail et problématique	1
Contribution	2
Organisation du mémoire	3

Première Partie:Etat de l'Art

Chapitre I : Les Annotations

I.1. Introduction	6
I.2. Définitions	6
I.3. Structure de l'annotation	7
I.3.1 L'objet Annotation	7
I.3.2. L'activité Annotation	11
I.4. Sémantique de l'annotation (objectifs)	12
I.4.1. Les objectifs selon C. Marshall [Marshall, 98]	12
I.4.2. Les objectifs selon J. Virbel [Veron, 97]	13
I.4.3. Les objectifs selon [Mille, 05].....	14
I.5. Catégorisation des annotations	14
I.5.1. Les catégories de l'objet annotation.....	15
I.5.1.1. L'annotation cognitive.....	15
I.5.1.2. L'annotation computationnelle.....	15
I.5.1.3. L'annotation sémantique	15
I.5.2. Les catégories de l'activité annotation	15
I.5.2.1.L'annotation manuelle.....	15
I.5.2.2. L'annotation semi-automatique.....	16
I.5.2.3. L'annotation automatique.....	16
I.6. Les outils d'annotations	16
I.7. L'annotation sémantique	18
I.7.1. Définitions.....	18
I.7.2. Les langages d'annotation sémantique.....	19
I.8. Conclusion	19

Chapitre II : Le Tagging Collaboratif

II.1. Introduction	21
II.2. Définitions	21
II.2.1. Le Tagging collaboratif.....	21
II.2.2. Le Tag	22
II.2.3. La Folksonomie	22
II.3. Structure d'une action du Tagging collaboratif	24
II.3.1. Structure tripartite de base	24
II.3.2. Structure tripartite avec liens inter-ressources et inter-utilisateurs.....	25
II.3.3. Structure quadripartite	26
II.4. Propriétés d'un système du Tagging collaboratif	27
II.5. Tagging vs Annotation	28
II.6. Etude des systèmes du Tagging collaboratif	28
II.6.1. Etude de la dynamique des systèmes du Tagging.....	29
II.6.2. Travaux sur la proposition de modèles et d'algorithmes de suggestion de tags et d'utilisateurs	30
II.6.3. Découverte de communauté.....	31
II.7. Tagging collaboratif et recherche d'information	31
II.8. Rapprocher les ontologies et les folksonomies	33
II.8.1. Les approches d'extraction de liens sémantiques entre tags.....	33
II.8.1.1. Analyse des réseaux sociaux appliquée aux folksonomies	33
II.8.1.2. Analyse de la dynamique des folksonomies	33
II.8.1.3. Clustering	34
II.8.2. Les approches basées sur les ontologies	34
II.8.2.1. Guider le Tagging à l'aide d'ontologies.....	34
II.8.2.2. Construire une ontologie de folksonomies.....	34
II.8.3. Exemples d'ontologies informatiques pour le Tagging	34
II.9. Les limites des systèmes du Tagging	36
II.10. Conclusion	37

Chapitre III : Le Profil Utilisateur

III.1. Introduction	38
III.2. Définition	38
III.3. Modélisation du Profil	39
III.4. Représentation du Profil	40
III.4.1. Représentation ensembliste ou vectorielle.....	40
III.4.2. Représentation sémantique	40
III.4.3. Représentation connexionniste	40

III.4.4. Représentation multidimensionnelle.....	40
III.4.5. Représentation hiérarchique.....	40
III.5. Dimensions d'un Profil utilisateur	41
III.6. Acquisition du Profil	42
III.6.1. Approche simpliste :	43
III.6.2. Approche dynamique :	43
III.6.3. Approche par apprentissage :	43
III.7. Modèle conceptuel du profil	43
III.8. Profil utilisateur dans le Web 2.0	44
III.8.1. Recherche à base de tags et profil utilisateur	45
III.8.2. Créer et enrichir le profil utilisateur en se basant sur les tags	45
III.9. Conclusion	47

Deuxième Partie

Une Approche de Filtrage de Tags à base du Profil utilisateur

Chapitre IV : Présentation de l'approche

IV.1. Introduction.....	49
IV.2. Motivations	50
IV.3 Principe général	50
IV.4 Présentation de l'approche	51
IV.4.1. Le modèle du profil utilisateur	53
IV.4.1.1. Représentation du profil	53
IV.4.1.1.1. La dimension personnelle.....	52
IV.4.1.1.2. La dimension centres d'intérêts.....	52
IV.4.1.1.3. La dimension expertise.....	52
IV.4.1.2. Construction du profil.....	55
IV.4.1.2.1. Construction de la dimension centres d'intérêts.....	54
- L'approche hybride.....	54
- L'algorithme Add-A-Tag adapté à notre approche hybride.....	56
IV.4.1.2.2. Construction de la dimension expertise.....	56
- Choix de l'ontologie.....	57
- Profondeurs des termes dans WordNet.....	58
IV.4.2. Pondération des tags à base du profil utilisateur	60
IV.4.2.1. Etude des variations de la formule de pondération.....	63
IV.4.3. Classement des tags et construction des descripteurs (filtrage).....	65
IV.5 Conclusion.....	66

Chapitre V : Tests et Evaluations

V.1. Introduction	67
V.2. Collection de test.....	67
V.3. Démarche d'évaluation	67
V.3.1. Phase de préparation de la collection.....	68
V.3.1.1. Elimination des tags ne figurant pas dans WordNet.....	68
V.3.1.2. Récupération des profondeurs des tags à partir de WordNet.....	68
V.3.1.3. Indexation du contenu textuel des pages web.....	69
V.3.1.3. Elimination des mots-clés de l'index ne figurant pas dans WordNet.....	71
V.3.2. Evaluation en utilisant un système de recherche d'information (SRI).....	71
V.3.2.1. Implémentation du moteur de recherche.....	71
V.3.2.2. Construction du vecteur idéal	72
V.3.2.3. Construction du vecteur popularité	73
V.3.2.4. Construction du vecteur poids	74
V.3.3. Comparaison des résultats	75
V.4. Architecture du système d'évaluation	76
V.5. Résultats et discussion.....	79
V.6. Conclusion.....	81
Conclusion Générale	
Synthèse.....	82
Résumé de la contribution.....	82
Perspectives.....	83
Bibliographie	
Annexes	

Liste des Figures

Fig 01 :	Représentation d'une annotation dans Annotea.	10
Fig 02 :	Exemple d'une annotation sémantique avec HTML-A.	19
Fig 03 :	Structure tripartite d'un ensemble d'actions du Tagging collaboratif.	25
Fig 04 :	Modèle conceptuel d'un système de Tagging [Marlow, 06].	26
Fig 05 :	Structure quadripartite d'un ensemble d'actions du Tagging.	27
Fig 06 :	Schéma du modèle structurel de CommonTag.	35
Fig 07 :	Processus de communication entre un client et un serveur MOAT.	35
Fig 08 :	Action du Tagging pour un contenu donné, [Passant,08].	36
Fig 09 :	Dimensions et sous-dimensions d'un modèle du profil.	42
Fig 10 :	Modèle conceptuel du profil utilisateur.	44
Fig 11 :	Principe général de l'approche.	51
Fig 12 :	Schéma global de l'approche.	52
Fig 13 :	Processus détaillé de l'approche.	53
Fig 14 :	Dimensions du Profil utilisateur défini dans l'approche.	55
Fig 15 :	Exemple de graphe construit avec la combinaison des deux approches	56
Fig 16 :	Comparaison des vecteurs intérêts construits avec les différentes.	56
Fig 17 :	Exemple de définition de terme dans WordNet.	59
Fig 18 :	Vecteurs d'intérêts des quatre utilisateurs.	62
Fig 19 :	Courbe illustrant les variations du poids en fonction d'expertise.	63
Fig 20 :	Courbe illustrant les variations du poids en fonction de distance.	64
Fig 21 :	Courbe illustrant les variations du poids en fonction de confiance.	65
Fig 22 :	Processus d'évaluation de l'approche.	68
Fig 23 :	Diagramme d'activité du processus de préparation des tags.	69
Fig 24 :	Diagramme d'activité du processus d'indexation	70
Fig 25 :	Diagramme d'activité du processus de recherche.	72
Fig 26 :	Construction du vecteur idéal (VI).	73
Fig 27 :	Construction du vecteur popularité (VP).	74
Fig 28 :	Construction du vecteur poids (VW).	75
Fig 29 :	Courbe de comparaison des vecteurs.	76
Fig 30 :	Architecture du système d'évaluation.	77
Fig 31 :	Diagramme des cas d'utilisation.	78
Fig 32 :	Diagramme des classes	79
Fig 33 :	Comparaison entre la recherche à base de la popularité et du poids.	80

Liste des Tableaux

Tableau 01 : Exemples de schéma de création d'annotations [Azouaou, 06]. . .	12
Tableau 02 : Classification des Outils d'annotations existants, [Azouaou, 05]...	17
Tableau 03 : Comparaison entre Tagging et Annotation	28
Tableau 04 : Problèmes dus à l'utilisation unique de la popularité.....	50
Tableau 05 : Exemple de calcul d'expertise utilisateur.....	59
Tableau 06 : Liste des tags associés à la ressource	62
Tableau 07 : Distances, Expertise et confiances des utilisateurs.	63
Tableau 08 : Classement des tags par ordre décroissant du poids.	65
Tableau 08 : Distances (Cosinus) entre les vecteurs VP, VW et VI.....	81

Liste des Formules

Formule 01 :Formule du calcul du rand d'une page web, [Xu, 08].	32
Formule 02 :Formule de création du profil, [Firan, 07].....	46
Formule 03 :Calcul du poids d'un tag basé sur l'ordre, [Huang, 08].....	47
Formule 04 :Formule du calcul de la capacité d'un tag, [Huang, 08].....	47
Formule 05 :Formule du calcul de l'Expertise.....	58
Formule 06 :Formule de pondération d'un tag à base du profil utilisateur.....	60
Formule 07 :Formule de la mesure cosinus, [Gerald, 05].....	60
Formule 08 :Formule du calcul de la distance utilisateur-ressource.....	61
Formule 09 :Formule du calcul de la confiance.....	61